The peptide-catalyzed stereospecific synthesis of tetroses: A possible model for prebiotic molecular evolution

Arthur L. Weber* and Sandra Pizzarello†‡

*Ames Research Center, SETI Institute, Moffet Field, CA 94035-1000; and †Department of Chemistry and Biochemistry, Arizona State University, Tempe, AZ 85018-1604

Edited by Stephen J. Benkovic, Pennsylvania State University, University Park, PA, and approved July 5, 2006 (received for review March 21, 2006)

Using a water-based prebiotic model of sugar synthesis involving glycolaldehyde self-condensation, we demonstrate that homochiral L-dipeptide catalysts lead to the stereospecific syntheses of tetroses. The asymmetric effect is largest for erythrose, which may reach a D-enantiomeric excess of >80% with L-Val-L-Val catalyst. Based on results obtained with various peptides, we propose a possible catalytic-reaction intermediate, consisting of an imidazolidinone ring formed between the two nitrogen atoms of the peptide catalyst and the C1 of one glycolaldehyde molecule. The study was motivated by the premise that exogenous material, such as the nonracemic amino acids found in meteorites, could have participated in the terrestrial evolution of molecular asymmetry by stereospecific catalysis. Because peptides might have formed readily on the early Earth, it is possible that their catalytic contribution was relevant in the prebiotic processes that preceded the onset of life.

prebiotic chemistry | chirality | catalysis

Although the chemical environments that led to life's origin on the Earth are unknown, our knowledge of current and extinct life processes allows us to postulate some possible working assumptions. For example, we may describe life as an autocatalytic chemical process where simple molecules, such as amino acids and nucleotides, assemble into larger proteins and nucleic acids that, in turn, control their metabolism and synthesis (1). In these terms, we may also identify key requirements that make such processes possible. One such prerequisite is the exclusive one-handedness of the chiral constituents of proteins and nucleotides (L-amino acids and D-sugars) that is essential to extant biopolymers' structure and function and whose origin is unknown.

Several general theories have been brought forward in regard to the origin of this terrestrial homochirality (2). Some propose that life first arose from a racemic environment and later evolved to chiral homogeneity; this "biotic" premise circumvents the ease with which prebiotic amino acids and sugars would have lost any incremental gain in asymmetry to racemization. Abiotic theories, on the other hand, propose that some asymmetry preceded the onset of life, overcame racemization, and contributed to homochirality. The finding in meteorites of amino acids resistant to racemization and displaying enantiomeric excesses (ee) that, if less extensive, have the same sign (L-) as in terrestrial amino acids (3, 4) supports, albeit does not prove, the latter hypothesis.

Apart from either hypothesis, because abiotic syntheses of chiral molecules would produce only racemic mixtures if unaided by catalysts and meteorites' composition shows that any preexisting ee available for the formation of biomolecules would not have been large (4), it is reasonable to assume that the development of chiral homogeneity was an evolutionary process that involved catalysis. This inference, in turn, leads to the question of whether the unique nonracemic amino acids of meteorites might have played a role in the terrestrial evolution of homochirality by transferring their asymmetry catalytically to other prebiotic building blocks, such as sugars.

Protein amino acids, small peptides, and amines are known catalysts that mediate ee amplification of the products in –C–C– bond-forming reactions (e.g., see refs. 5 and 6). In particular, the intermolecular aldol condensation of ketones and aldehydes under stereospecific catalysis by proline and its derivatives has been studied in detail (7–12). These studies have offered insights into the possible mimicking of bioenzymatic processes by simple organic molecules (13, 14) and comprise an ample, well reviewed literature (15). However, besides indications that some catalyzed aldol reactions could tolerate small (<4 vol %) amounts of water in the reaction medium (16) and proceed with certain organocatalysts in combination with buffers and surfactants (17), all methodologies in these first experiments required organic solvents, such as acetone, DMSO, THF, and dimethylformamide, which are not consistent with the realistic expectations of an exclusively water-based prebiotic environment.

Recently (18), we studied a prebiotic model of sugar synthesis in which nonracemic alanine and 2-ethylglycine (isovaline), chosen as representative of protein and meteoritic amino acids respectively, were both found to act as asymmetric catalysts in water as well. The aldol condensation of glycolaldehyde conducted in ammonium acetate buffer at no more than 50°C lead to the synthesis of the tetrose sugars with significant ee. The configuration of both threose and erythrose products were affected at C-2, through the formation of an imine intermediate, and the largest ee (12%) was obtained for D-threose with L-isovaline catalyst. Surprisingly, proline did not have any catalytic effect under these conditions.

In subsequent studies, Córdova *et al.* (19) used a proline catalyst and *N*,*N*-dimethylformamide solvent for the synthesis of nonracemic hexoses by the sequential aldol condensation of glycolaldehyde. Also, Zou *et al.* (20) demonstrated that dipeptides in DMSO with the addition of 2–8% water could catalyze the aldol condensation of several ketone and aldehyde alkyl derivatives, including a single example of peptide-catalyzed self-condensation of glycolaldehyde in water with modest ee. Finally, Dziedzic *et al.* (21) have reported that peptide catalysts could also provide stereoselectivity for these reactions in phosphate buffers if surfactants were added.

After a brief earlier communication (22), we report here the asymmetric effect of a series of homochiral dipeptide catalysts on the stereospecific synthesis of tetroses via glycolaldehyde self-condensation. As in the previous study of amino acid catalysts, these experiments were carried out under realistic aqueous

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviation: ee, enantiomeric excess(es).

‡To whom correspondence should be addressed. E-mail: pizzar@asu.edu.

Table 1. Stereospecific synthesis of tetroses by the condensation of glycolaldehyde using dipeptide catalysts containing a β-branched amino acid

Peptide catalyst	Temp, °C (time, h)	Buffer	pH	D-thr, % ee	D-ery, % ee	% yield thr + ery*	ery/thr
L-Val-Gly	50 (3)	0.25 M TEAA	5.4	2	−11	50	1
Gly-L-Val	50 (3)	0.25 M TEAA	5.4	2	53	20	1.2
Gly-L-Ile	50 (3)	0.25 M TEAA	5.4	3	53	15	1.5
L-Ala-L-Val	50 (3)	0.25 M TEAA	5.4	0	40	17	1.2
L-Ala-L-Ile	50 (3)	0.25 M TEAA	5.4	2	38	5	1
L-Val-L-Val	50 (3)	0.25 M TEAA	5.4	−1	59	25	1
L-Ile-L-Val	50 (3)	0.25 M TEAA	5.4	−1	61	25	1
L-Val-L-Val	50 (3)	0.05 M TEAA	5.4	−2	75	22	1
L-Val-L-Val	25 (18)	0.25 M TEAA	5.4	−1	72	5	1.5
L-Val-L-Val	25 (18)	0.05 M NaAc	5.4	−7	80	23	1.5
L-Val-L-Val	25 (18)	0.05 M CaAc	5.4	−10	78.5	20	1.5
L-Val-L-Val	25 (18)	0.05 M NaAc	4.9	−1	82	12	1.5
L-Val-L-Val	0 (32)	0.05 M NaAc	5.4	1.5	13	29	1.7

thr, threose; ery, erythrose; TEAA, triethylammonium acetate; NaAc, sodium acetate; CaAc, calcium acetate.
*Yield of eythrose plus threose as percent of residual starting glycolaldehyde and other dimeric products of condensation.

conditions to assess any relevance the reactions may have for prebiotic chemistry.

Results and Discussion

The results, determined by GC-MS analysis, of the reaction products are given in Tables 1 and 2 and summarized in Fig. 1, which offers an overview of the D-ee obtained for erythrose and threose synthesized from glycolaldehyde in the presence of various dipeptide catalysts as well as reference to those found previously upon catalysis by a single amino acid. It is apparent that the asymmetric effect of most of the dipeptides on the glycolaldehyde aldol condensation products differs from that observed with amino acid catalysts in both the extent of the stereospecific effect and the tetrose affected. In these peptide reactions, in fact, the main recipient of the asymmetric effect is erythrose, whereas threose is affected only partially or not at all.

Between the catalysts analyzed, the ee was found to be higher for erythrose produced by dipeptides with a β-branched alkyl chain near the C terminus, to increase at reduced temperature with higher reaction time up to 25°C and to decrease again at

Table 2. Stereospecific synthesis of tetroses by the condensation of glycolaldehyde using peptide catalysts without a β-branched amino acid

Peptide catalyst	Temp, °C (time, h)	Buffer	pH	D-thr, % ee	D-ery, % ee	% yield thr + ery*	ery/thr
L-Ala-Gly	50 (3)	0.25 M TEAA	5.4	−16	2	50	1
L-Ala-Gly-Gly	50 (3)	0.25 M TEAA	5.4	−16	−3	15	1
L-Phe-Gly	50 (3)	0.25 M TEAA	5.4	−5	7	63	1
Gly-L-Pro	50 (3)	0.25 M TEAA	5.4	0	0	17	0.8
Gly-L-Ala	50 (3)	0.25 M TEAA	5.4	0	43	10	1.8
L-Ala-L-Ala	50 (3)	0.25 M TEAA	5.4	−7	33	45	1.5
L-Ala-L-Ala	50 (1.5)	0.25 M TEAA	5.4	−9	33	10	2
L-Ala-L-Ala	0 (96)	0.25 M TEAA	5.4	0	0	50	0.8
Tri-L-Ala	50 (3)	0.25 M TEAA	5.4	−13	22	17	0.8
Tetra-L-Ala	50 (3)	0.25 M TEAA	5.4	−12	17	5	1.2
L-Leu-L-Ala	50 (3)	0.25 M TEAA	5.4	−4	35	38	1
L-Ala-L-Phe	50 (3)	0.25 M TEAA	5.4	1	31	5	1
L-Ala-L-Glu	50 (3)	0.25 M TEAA	5.4	32	32	11	1
L-Ala-L-Leu-L-Ala	50 (3)	0.25 M TEAA	5.4	−13	33	13	1
L-Ala-L-His	50 (3)	0.25 M TEAA	5.4	ND	ND	Trace	ND
D-Ala-L-Ala-L-Ala	50 (3)	0.25 M TEAA	5.4	9	7	15	1.2
L-Ala-L-Ala	50 (3)	0.05 M NaAc	5.4	−4	17	34	1
D-Ala-D-Ala	50 (3)	0.05 M NaAc	5.4	2	−13	60	0.8
D-Ala-D-Ala	25 (18)	0.05 M NaAc	5.4	0.5	−11	34	1.2
L-Glu-L-Glu	50 (3)	0.25 M TEAA	5.4	−1.5	34	34	1.2
L-Leu-L-Leu†	50 (3)	0.25 M TEAA	5.4	0	0	<5	1
β-Ala-L-Ala	50 (3)	0.25 M TEAA	5.4	0	10	25	1
Ac-(L-Ala)$_3$	50 (3)	0.25 M TEAA	5.4	ND	ND	None	ND

thr, threose; ery, erythrose; TEAA, triethylammonium acetate; NaAc, sodium acetate; ND, not determined because of insufficient yield.
*Yield of eythrose plus threose as percent of residual starting glycolaldehyde and other dimeric products of condensation.
†Precipitate formed during reaction.

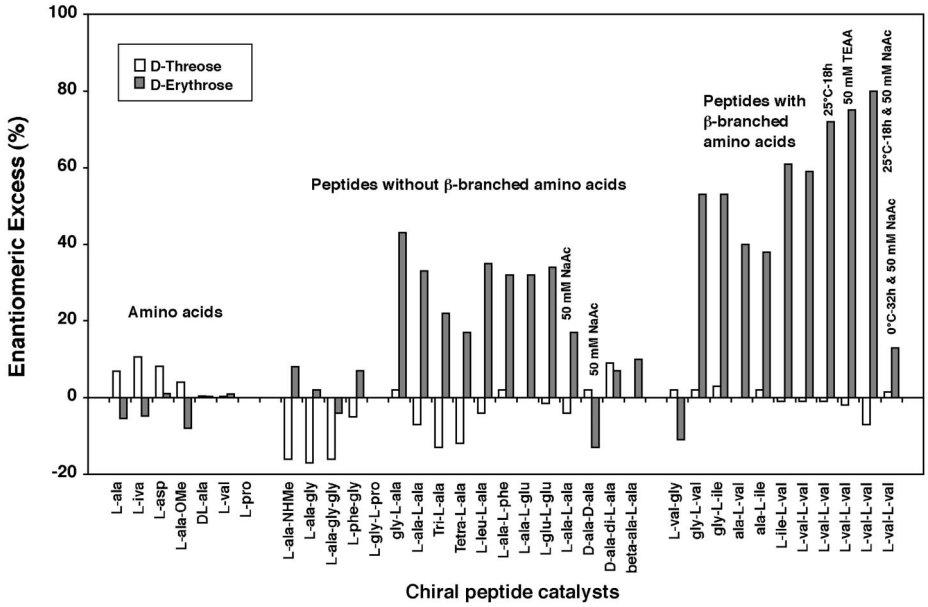
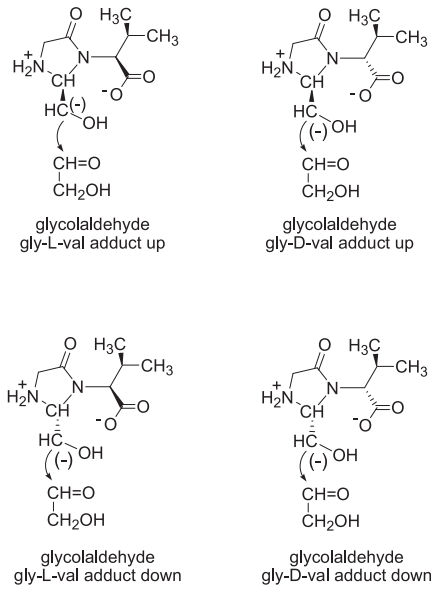


Fig. 1. Summary plot of the distribution of ee in erythrose and threose synthesized by the self-condensation of glycolaldehyde in the presence of various homochiral dipeptide catalysts.

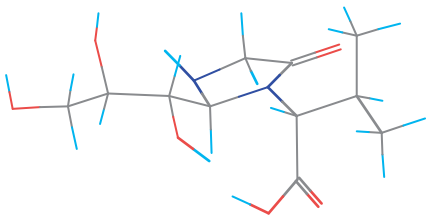
0°C. The effect of β-branching on enantioselectivity points to the importance of the alkyl side chain to the reaction and is confirmed by the results obtained with valine peptides. L-Val-L-Val catalyst gave the highest ee at 25°C after 18 h, and the ratio of D-erythrose/L-erythrose/D-threose/L-threose in the sample was found to be ≈10:1:3:3. The selective production of D-erythrose by several catalysts suggests that the reaction mechanism should favor the synthesis of this sugar at the expense of the L-erythrose, and the D- and L-threose enantiomers. In the case of L-ala, the dipeptide gave a higher D-erythrose enrichment than the homochiral tri- and tetrapeptides.

Product yields shown in Tables 1 and 2 are the combined chromatographic peak areas determined for erythrose and threose expressed as percent of total peak area, which also included starting material and unknowns. For reactions using dipeptide catalysts with a β-branched amino acid (Table 1), where the larger erythrose ee were produced, yields were found to be similar at 25°C and 50°C and have an average of 21% when the lowest and highest values of the series are disregarded. This relatively low recovery is not surprising, because we examined the early part of the glycolaldehyde condensation reaction to measure its stereospecificity (and relative rates) with minimum interference from possible aldo–keto isomerization side reactions, and it is plausible that the same steric hindrance favoring high stereospecificity for the branched dipeptide intermediates (see below) also hampered tetrose production. The dipeptide catalytic effect on glycolaldehyde self-aldolization can be reasonably implied by the increased tetrose yields of catalyzed versus uncatalyzed reactions, the full recovery of the dipeptide obtained after the reaction, and the stereoselective increase in the tetrose syntheses conducted with different dipeptide catalysts. The above inference is also supported by the fact that amines are well known catalysts of the aldol condensation reaction (e.g., refs. 1 and 15). It will be useful to define the as yet unknown products of this reaction, which appear, by their mass spectra, to be dimeric products of glycolaldehyde and tetroses (see *Materials and Methods*). They also display ee.

The divergence between amino acid and dipeptide data as well as the variation of ee yields between dipeptide catalysts could be accounted for by the formation of an imidazolidinone reaction intermediate between the dipeptide catalyst and the first glycolaldehyde adduct, as shown in Scheme 1 for Gly-L-Val and Gly-D-Val catalysts. For a racemic peptide, this ring formation

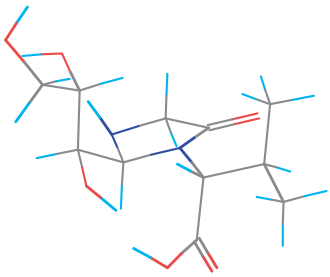


Scheme 1.



Scheme 2.

BIOCHEMISTRY



Scheme 3.

leads to two asymmetric centers, giving four possible diastereomeric intermediates.

Unlike in amino acid catalysis, this intermediate would not benefit from an imine resonance to stabilize the nucleophile involved in glycolaldehyde condensation. Rather, it appears to rely on the positioning of the peptide carboxyl group near the hydrogen of the C-2 hydroxyl and, possibly, on electrostatic inductive and field effects. Although it is difficult to gauge the role of water in such catalytic mechanisms, and Scheme 1 is left open to further interpretation, it is consistent with several observations. For example, the ee increase with steric constraint of the catalyst's free carboxyl group, the highest ee yielded by C-terminal valine peptides, where the alkyl group would provide further rigidity to the catalyst's free carboxyl, and the marked decrease in stereospecific catalytic effect with the increasing length of the linear ala-peptides, where the carboxyl is further and further removed from the reaction center.

The mechanistic participation of the peptide carboxylate also appears to be directly supported by the dominance of the second amino acid residue in determining catalytic stereoselectivity, as seen in (*i*) the discrepancy of the L-Val-Gly versus Gly-L-Val and L-Ala-L-Ala versus L-Ala-Gly data, (*ii*) the synthesis of racemic tetroses by Gly-L-Pro, a peptide that lacks the hydrogen needed to form an imidazolidinone, and (*iii*) the low stereoselectivity of the 0°C reactions, suggesting a minimum temperature for such effect, which could be explained by a kinetic barrier inhibiting formation and breakdown of stereoselective imidazolidinone-ring intermediates.

Models of homochiral intermediates containing the final D- and L-tetrose adducts appear to confirm the above assumption. As shown below, the adduct of D-erythrose (Scheme 2) points the tetrose sugar away from the imidazolidinone ring and is sterically favored compared with that of L-erythrose (Scheme 3), which would put the sugar chain close to the ring's alkyl substituents.

Imidazolidinone rings between peptide N-terminal and simple reactive aldehyde and ketones have been synthesized by using solution chemistry (23) and are known to adopt preferential stereochemical orientations (24). That the same or, possibly, similar rings could form in aqueous solution to give intermediates with preferred stereochemistry as well as the scope of the ee obtained in this medium are of interest in the general context of organocatalysis.

From the perspective of prebiotic chemistry, the possibility of an asymmetric catalytic effect of the observed magnitude between biomolecule precursors is appealing, and it is interesting to note that D-erythrose has the same molecular structure at the penultimate and third to the last carbons as D-ribose. The formation of dipeptides could reasonably be considered a simple evolutionary step for prebiotic amino acids to undertake on the early Earth, for example, by the facile reaction between amino acids and carbonyl sulfide near volcanic sources (25). On the other hand, the step should also be expected to be stochastic in the distribution of homochiral and racemic peptide products. For homochiral and prebiotically significant dipeptides, therefore, we need to consider a possible asymmetric induction by other material available in early Earth environments. To this end, the L-ee of meteoritic amino acids, which were validated unequivocally by isotopic analyses (26), offer an indication that such asymmetry was achievable in abiotic chemistry and may have been wide-ranging and made available to the early Earth during impact delivery of asteroidal and cometary material (ref. 27 and references therein).

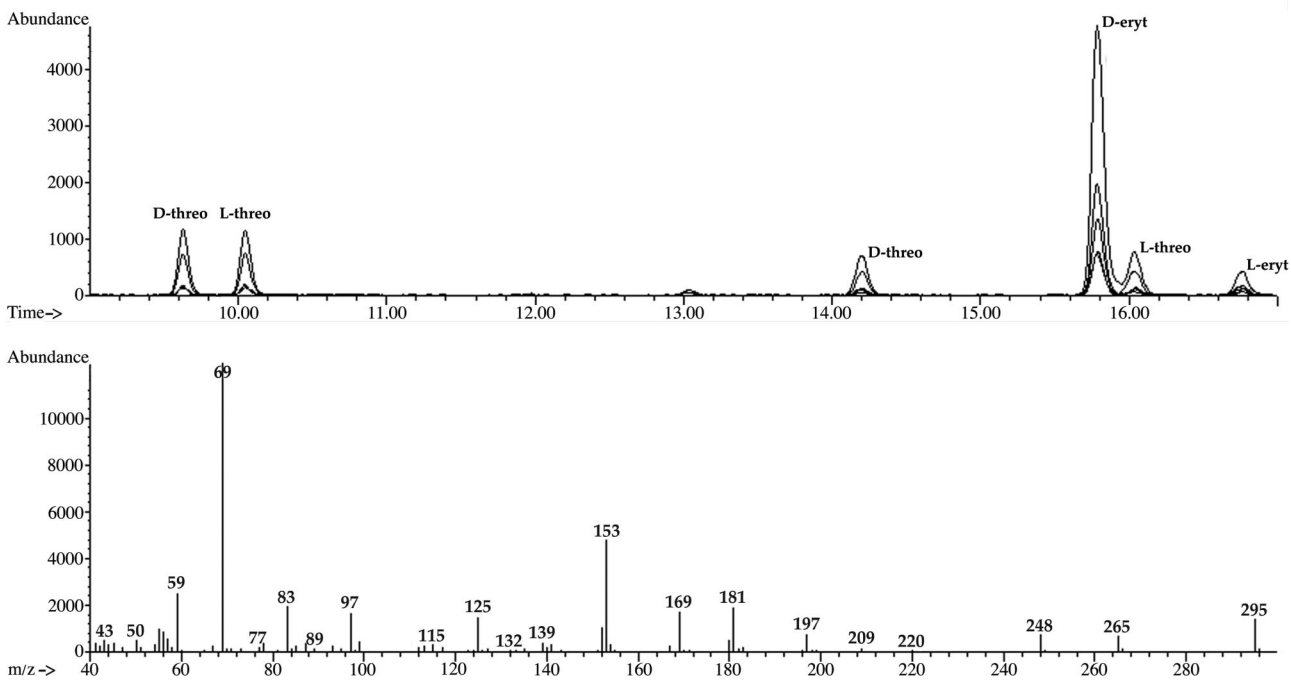


Fig. 2. GC-MS analysis of erythrose and tetrose produced by glycolaldehyde self-condensation with L-Val-L-Val catalyst. Single ion traces at *m/z* 295, 181, 248, and 295 of the sugar *O*-TFA derivatives.

Materials and Methods

Peptides and other chemicals were obtained from commercial suppliers. Unless indicated otherwise in the tables and figures, the aqueous catalyzed reactions used 120 μmol of glycolaldehyde that was reacted in the presence of 30 μmol of peptide catalyst in a 0.25 M triethylammonium acetate buffer, pH 5.5 (1.2 ml total volume), at 50°C, for 3 h *in vacuo* in sealed ampules. The reaction products were kept frozen at −80°C until analyzed. Control reactions of D-erythrose in the absence and presence of amino acid or peptide catalysts gave <1% racemization at C-2 (i.e., isomerization to D-threose). Reaction of glycolaldehyde in the absence of catalysts produced no ee of the tetrose products, whose yield was ≈1%.

For analysis of the reaction products, the reaction solution was brought to room temperature, and a one-third aliquot was applied to a column filled with ≈12 ml of cation-exchange resin in hydrogen form (200–400 mesh, analytical grade; Bio-Rad, Hercules, CA). The column was eluted first with water and then with 2 M ammonium hydroxide. Both eluates were then dried by rotary evaporator at 30°C. The dry water-eluate residue was reacted with trifluoroacetic anhydride at 25°C to give the *O*-TFA derivatives of the sugars, whereas the ammonium hydroxide residue was reacted first with 2-propanol (≈2 M HCl) for 24 h at 25°C and, after drying of the alcohol, with trifluoroacetic anhydride, as above. Unused portions of the reaction solution were immediately frozen for repeat analyses.

GC separation was performed with a capillary column coated with a bonded cyclodextrin phase (Chirasil-Dex CB; Chrompack; 25 m × 0.25 mm, 0.25-μm thickness). The oven temperature program was 70°C initial for 5 min, to 85°C at 1°C·min^{-1}, to 200°C at 4°C·min^{-1}, and 45 min of final temperature hold. Four peaks were obtained for each erythrose and threose, corresponding to the enantiomers of the two possible furanose structures of the aldotetroses, the noncyclic tetroses gave very small peaks, and erythulose enantiomers were only partially separated but allowed the detection of large ee, if present. Also observed were the peaks of unreacted glycolaldehyde dimers as well as those of later eluting compounds having mass spectra corresponding to glycolaldehyde-tetrose dimers. Of these, some showed ee comparable with those of erythrose, but their identity could not be confirmed for the absence of standards.

D-ee (D% − L%) give mean values based on n integrated ion intensities, where $n \geq 6$ and included multiple specific ions selected from multiple runs. Fig. 2 gives an example of the GC-MS separation of threose and erythrose enantiomers under these conditions for a L-Val-L-Val catalyst sample and the mass spectrum of the D-erythrose peak.

This work was supported by National Aeronautics and Space Administration Exobiology and Evolutionary Biology Program grants (to A.L.W. and S.P.).

1. Weber, A. L. (2001) *Orig. Life Evol. Biosph.* **31,** 71–86.
2. Cronin, J. R. & Reiss, J. (2005) in *Lectures in Astrobiology*, ed. Gargaud, M. (Springer, Berlin), Vol. 1, pp 473–514.
3. Cronin, J. R. & Pizzarello, S. (1997) *Science* **275,** 951–955.
4. Pizzarello, S., Zolensky, M. & Turk, K. A. (2003) *Geochim. Cosmochim. Acta* **67,** 1589–1595.
5. Jacobsen, N. E., Pfaltz, A. & Yamamoto, H., eds. (1999) *Comprehensive Asymmetric Catalysis* (Springer, New York).
6. Shibasaki, M., Yoshikawa, N. & Matsunaga, S. (2004) in *Comprehensive Asymmetric Catalysis*, eds. Jacobsen N. E., Pfaltz A. & Yamamoto H. (Springer, New York), Suppl. 1, pp. 135–142.
7. List, B., Lerner, R. A. & Barbas, C. F., III (2000) *J. Am. Chem. Soc.* **122,** 2395–2396.
8. Notz, W. & List, B. (2000) *J. Am. Chem. Soc.* **122,** 7386–7387.
9. List, B., Porjarliev, P. & Castello, C. (2001) *Org. Lett.* **3,** 573–575.
10. Córdova, A., Notz, W. & Barbas, C. F., III (2002) *J. Org. Chem.* **67,** 301–303.
11. Northrup, A. B. & McMillan, D. W. C. (2002) *J. Am. Chem. Soc.* **124,** 6798–6799.
12. Tung, Z., Jiang, F., Yu, L.-T., Cui, X., Gong, L.-Z., Mi, A.-Q., Jiang, Y.-Z. & Wu, Y.-D. (2003) *J. Am. Chem. Soc.* **125,** 5262–5263.
13. Gröger, H. & Wilken, J. (2001) *Angew. Chem. Int. Ed.* **40,** 529–532.
14. Movassaghi, M. & Jacobsen, E. N. (2002) *Science* **298,** 1904–1905.
15. List, B. (2002) *Tetrahedron* **58,** 5573–5590.
16. Sakthivel, W., Notz, W., Bui, T. & Barbas, C. F., III (2001) *J. Am. Chem. Soc.* **123,** 5260–5267.
17. Córdova, A., Notz, W. & Barbas, C. F., III (2002) *Chem. Commun.* **24,** 3024–3025.
18. Pizzarello, S. & Weber, A. L. (2004) *Science* **303,** 1151.
19. Córdova, A., Engqvist, M., Ibrahem, I., Casas, J. & Sundén, H. (2005) *Chem. Commun.* **2005,** 2047–2049.
20. Zou, W., Ibrahem, I., Dziedzic, P., Sundén, H. & Córdova, A. (2005) *Chem. Commun.* **2005,** 4946–4948.
21. Dziedzic, P., Zou, W., Hafren, J. & Córdova, A. (2006) *Org. Biol. Chem.* **4,** 38–40.
22. Pizzarello, S. & Weber, A. L. (2005) *Provisional Application to U.S. Patent Office*, filing no. 60/679,852.
23. Rinnova, M., Nefzi, A. & Houghten, R. A. (2002) *Tetrahedron Lett.* **43,** 2343–2346.
24. Roscic, M., Eklund, R., Nordmark, E.-L., Horvat, S. & Widmalm, G. (2004) *Eur. J. Org. Chem.* **2004,** 4641–4647.
25. Lemam, L., Orgel, L. & Gadhiri, M. R. (2004) *Science* **306,** 283–286.
26. Pizzarello, S. & Huang, Y. (2005) *Geochim. Cosmochim. Acta* **69,** 599–605.
27. Pizzarello, S. (2004) *Origins Life Evol. Biosph.* **34**, 25–34.